MINAR International Journal of Applied Sciences and Technology
ISSN: 2717-8234
Article type: Research Article

Received: 12/01/2023 **Accepted:** 15/02/2023 **Published:** 01/03/2023

CONSTRUCTING THE GENOME OF SARS-COV-2: AN ESSENTIAL STEP TOWARDUNDERSTANDINGTHE COVID-19 OUTBREAK THROUGH PHYLOGENETIC TREES

Nadia Moqbel Hassan ALZUBAYDI[1]

Computer Department, Engineering College, Mustansiriyah University, Baghdad, Iraq

Bashar Talib AL-NUAIMI[2]

Computer Department, Science College, Diyala University, Diyala, Iraq

Abstract

This paper discusses the emergence of (SARS-CoV-2) is a global concern due to its ability to cause coronavirus disease 2019 (COVID-19). In order better understand this virus and its evolutionary history, a phylogenetic tree was constructed using the Maximum Likelihood Estimate method and a Multiple Sequence Alignment of SARS-CoV-2 and other coronavirus species. This tree will be used to help elucidate the relationship between various coronaviruses and the spread of SARS-CoV-2. This paper made use of 46 SARS-CoV-2 isolates together with two other coronavirus species (Alpha & Beta) retrieved from the GenBank database (National Center for Biotechnology Information) for this exploration. Results indicated that the Maximum Likelihood Estimate approach was successful in producing a reliable phylogenetic tree that displayed the evolutionary history between SARS-CoV-2 and other coronavirus species, as well as delineating the genetic diversity within SARS-CoV-2.

Keywords: SARS-CoV-2 Genome, COVID-19, Maximum-Likelihood, Bioinformatics Tools, Phylogenetic Tree Construction.

http://dx.doi.org/10.47832/2717-8234.14.1

[1] nadiahasan@uomustansiriyah.edu.iq, https://orcid.org /0000-0002-6048-7223

[2] bashartalib6@gmail.com

Introduction

An in-depth analysis of the SARS-CoV-2 virus has been conducted with the aim of understanding its evolutionary history. By constructing a phylogenetic tree, scientists have been able to trace the detailed evolution of the virus from its origins to the present day [1]. This study provides an important insight into the spread of the virus, its mutations, and its potential for future mutation [2]. Using genetic sequencing techniques, a clear picture of the virus' genetic makeup has been developed. This data has been used to construct a phylogenetic tree that describes the relationship between different SARS-CoV-2 viruses and provides valuable information about its evolution [3]. The phylogenetic tree has enabled scientists to identify genetic variants and understand the virus' ability to evolve and adapt to its environment [4]. This study provides a comprehensive overview of the virus' evolution and will help researchers in the development of effective treatments and vaccines.

The novel coronavirus pandemic (COVID-19) has changed the world in many ways. Scientists are trying to understand the evolution of the virus that causes this disease, severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) [5] [6]. One way to do this is through phylogenetic tree construction [7] [8]. Phylogenetic tree construction is a process used to infer the evolutionary history of a group of related organisms (or viruses) by analyzing their genetic sequences. By constructing a phylogenetic tree, researchers can better understand the evolutionary patterns of SARS-CoV-2 and identify which areas of the world have been most affected by this virus [9]. This information can be used to inform public health strategies and vaccine development [10]. In this paper, we will discuss the basics of phylogenetic tree construction, explain how it is used to study the SARS-CoV-2 virus, and discuss the implications of this paper.

Phylogenetic tree construction involves the analysis of genetic data to create a visual representation of the evolutionary relationships among various organisms. The tree shows how closely related different organisms are and how they have evolved over time [11]. By studying the tree, researchers can gain insights into the transmission and spread of SARS-CoV-2, and identify the genetic changes that have occurred since its emergence.

The tree is then used to estimate how these species are related, this can help us to better understand the virus and its behavior, as well as inform strategies to prevent its spread [12]. Phylogenetic tree construction can be used to follow the development of novel virus strains, and to detect mutations that may be significant for its progression. Maximum likelihood approaches can be used to construct phylogenetic trees, and this approach can be used to examine SARS-CoV-2's evolution. [13].

MATERIAL AND METHODS

Data collection

NCBI's (National Center for Biotechnology Information) database was used to divide the entire genome from gene bank of (46) into (18) Alpha and (28) Beta coronaviruses [14].

Genomic information is a type of data that pertains to the DNA or genetic

material of an organism. It can include information such as accession numbers, definitions, genome size in base pairs (bp), paper titles, host information [15], and country of origin some of them are shown in the Table (1).

Phylogenetic trees were constructed from complete genome sequences by first aligning the sequences and then analyzing them to determine their relatedness [16]. Alignment of the sequences has been made using multiple sequence alignment, such as R language, to find regions of similarity among the sequences [17]. Once the sequences have been aligned, phylogenetic analysis can be performed using a number of methods [18], such as maximum likelihood. This method uses to infer the evolutionary relationships among the sequences [19]. Once the phylogenetic tree has been built, it can then be used to infer the evolutionary history of the species and identify relationships between them as shown in Figure1.

Table 1: Provides details on the genomes of SARS-COV-2 and two types of Coronaviruses (Alpha and Beta).

ACCESSION	DEFINITION	Genome size (bp)	Host	Country	Lineage
MW924112.1	Alphacoronavirus	28761	zoonotic bat	Korea	Alpha
MH817484.1	Feline alphacoronavirus	29137	Felis catus	Brazil	Alpha
OK287352.1	Alphacoronavirus	28435	avian species	Yunnan, China	Alpha
OP646598.1	Severe acute respiratory syndrome coronavirus 2	29684	birds and mammals	China, United Kingdom, Germany, Jordan, South Africa, Saudi Arabia, United Arab Emirates	Beta Sarbecovirus
NC_045512	SARS-CoV-2	29903	Homo sapiens	China	Beta Sarbecovirus
MW626421	Severe acute respiratory syndrome coronavirus 2	29802	Bat, Civet, Homo sapiens	Global	Beta Sarbecovirus

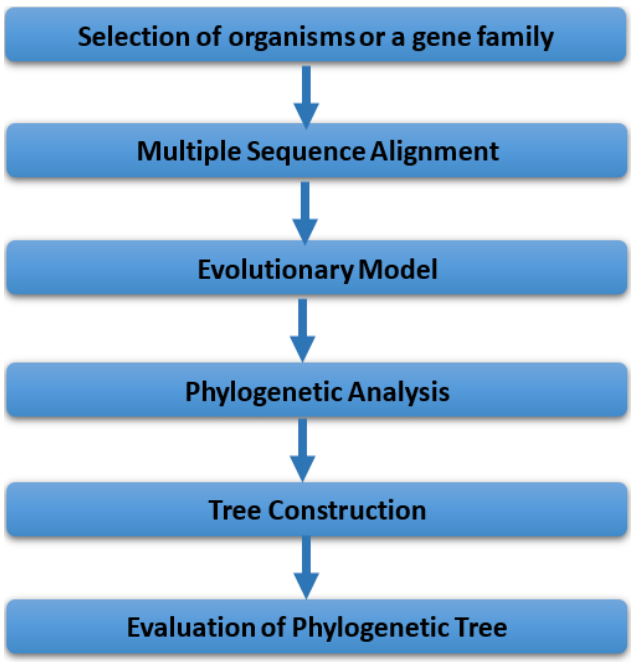

Figure 1: General proceedings for building a Coronavirus phylogenetic tree.

This step involves selecting the organisms or gene family to be included in the phylogenetic analysis [20]. This selection is important as it will determine the type and amount of data that will be used in the analysis, The genomes of alphacoronavirus and betacoronavirus were selected in this paper.

The Coronavirus and SARS-CoV-2 molecular data selection from the NCBI GenBank database provides access to genetic sequence data for the novel coronavirus SARS-CoV-2 and other coronaviruses. It includes the complete genomic sequences for these viruses through:

1. Navigate to the NCBI GenBank homepage (https://www.ncbi.nlm.nih.gov/genbank/) and select the "Nucleotide" database.
2. Enter the term "coronavirus" in the search box, and select the "All Fields" option.
3. Use the "Organism" and "Gene Family" filters to narrow down your search results to the desired organisms or gene family.
4. Select the "Send to" option for the desired search results, and choose "File", "FASTA" or "GenBank" as the file format.
5. The downloaded file can then be used to analyze the selected organisms or gene family.

The Coronavirus and SARS-CoV-2 molecular data Selection from the NCBI GenBank database is a collection of genetic sequences related to coronaviruses and SARS-CoV-2, the virus responsible for the COVID-19 pandemic. It includes genome sequences and more specialized data such as cDNA sequences, which are transcripts of mRNA sequences that can be used to study gene expression. The data is freely available to the public and is updated regularly [21].

Multiple Sequence Alignment

This step involves aligning the sequences of the selected organisms or gene family to identify similarities and differences in the genetic sequences. This can be done using various software packages such as ClustalW, MUSCLE, and T-Coffee. The goal of MSA is to find the most optimal alignment of the sequences that maximizes the similarity among them. to optimize the alignment of the sequence [22].

And here you will come the process of identifying regions of conserved gene order within the genomes of two different species, it is the "Synteny". It is a term used to describe the arrangement of genes and/or other DNA sequences in the genome that are located on the same chromosome. Before Multiple Sequence Alignment (MSA) used to construct a phylogenetic tree, these sequences must be compared to each other to determine the degree of similarity and evolutionary relatedness [23]. By looking at the synteny of the genes, it is possible to locate conserved blocks of DNA, which can provide evidence of common ancestry. This information can then be used to guide the MSA process, allowing the most accurate phylogenetic tree to be constructed. The FindSynteny() function takes the genomes as input, and it then identifies regions of similarity and rearrangements between the genomes. This is done by comparing and aligning the genomes. The output of the function is a graph showing the synteny blocks, which are regions of conserved gene order between two genomes [24] [25]. The graph also shows any rearrangements or inversions between the two genomes.

Generally, synteny matching takes place when two genomes have the same sections. We were able to get the first representation of the synteny of Alpha coronavirus with SARS-CoV-2 Wuhan-Hu-1 (NC_045512) outgroup, which is shown in Figure 2, and Beta coronavirus genomes, shown in Figure 3. We noticed blocks of similar sequences between them. These genomes demonstrate a blend of sequence likeness and divergence, with relatively few recombination events.

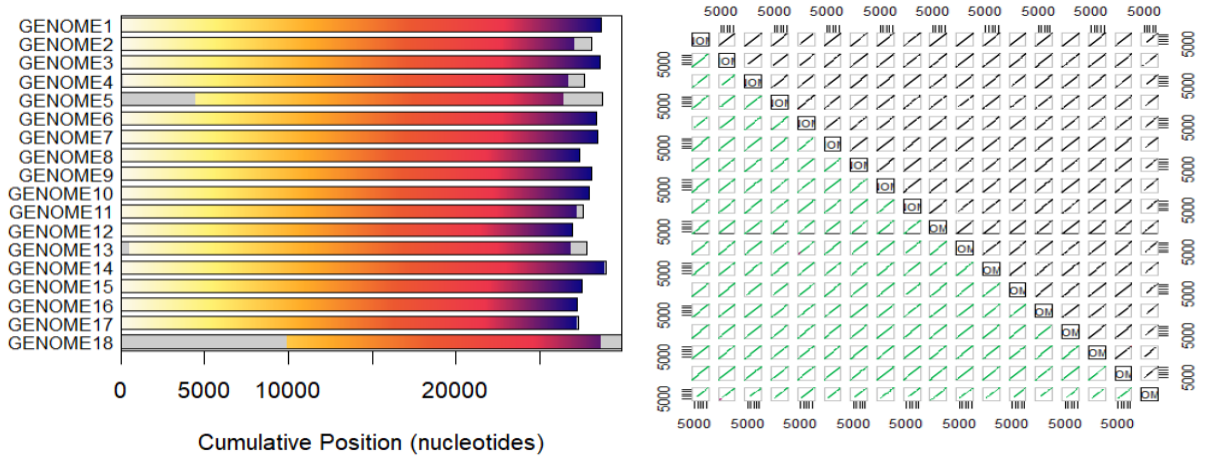

Figure.2. (a) A representation of Alpha coronavirus sequences through adjacent pairs synteny blocks. The blocks are color-coded according to the region they share with the first genome. If the regions are the same, the blocks are the same hue. If they don't share the same region, the block is colored grey. The size of the blocks is proportional to the length of the region they share. This type of representation is useful for visualizing the similarities and differences between different genomes. (b) Dot plot depicting the homologous regions among five Influenza virus A genomes. The dot plots show the similarities between the Alphacoronaviruses. The dot plots also reveal several regions of divergence between the genomes, which likely represent regions of antigenic variability.

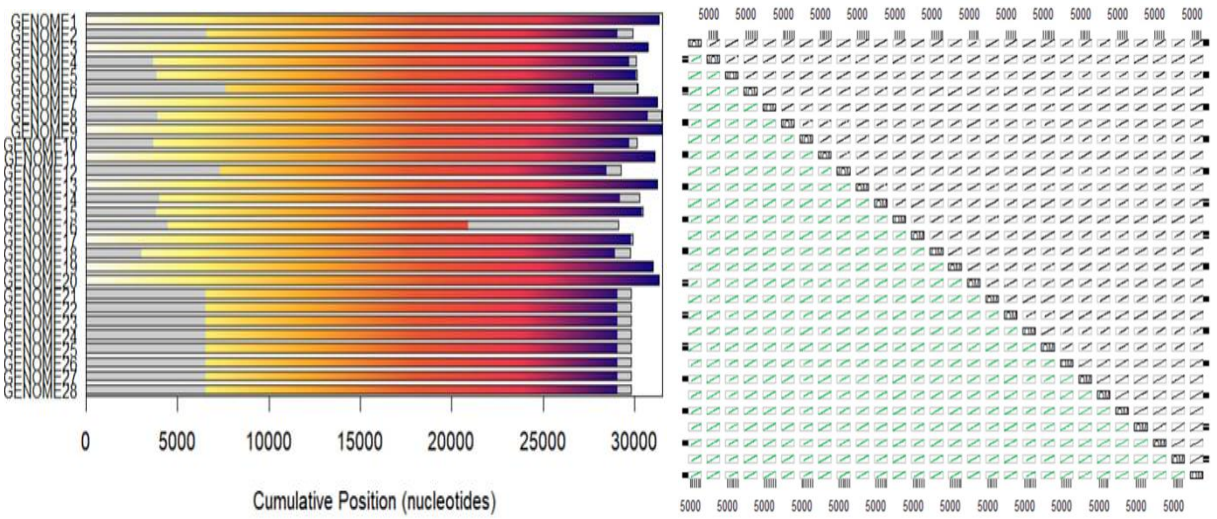

Figure.3. (a) Addition to the synteny maps, the alignment of the two genomes provides a more precise view of their similarity. The alignment can be visualized as a dot plot (b) Dot plot each nucleotide (or amino acid) is represented as a single point in a graph, and the points are connected by black lines if they are in the same location in both genomes. This visualization helps to identify regions of significant similarity, as well as regions of significant difference. The dot plot also allows for the identification of recombinant regions, which are regions of the genome that have been acquired from a different

Evolutionary model

This step involves selecting a model of evolution that best fits the data. This model will be used to generate a hypothesis of the evolutionary history of the organisms or gene family. Evolutionary tree model selection is an important step in constructing phylogenetic trees for coronaviruses. It involves selecting the appropriate evolutionary tree model to accurately represent the evolutionary relationships between the different coronaviruses. This model selection is based on the alignment of the nucleotide or amino acid sequences of the different coronaviruses. The evolutionary tree model should be chosen based on the characteristics of the sequence data, such as the number of sequence differences, the position of the differences, and the rate of evolution. Once the appropriate evolutionary tree model is chosen which is the GTR-GAMMA model, the phylogenetic tree can be constructed using a variety of software tools that are available. The resulting phylogenetic tree can then be used to identify the evolutionary relationships between the different coronaviruses, which can provide important information about the spread and evolution of the virus [26] [27].

The Maximum Likelihood Estimate (MLE) uses a tree-like structure to represent the evolutionary relationships among the species and calculates the likelihood of the observed data given the model. The model selection is based on the maximization of the likelihood of the observed data given the tree topology. This technique has been used to construct phylogenetic trees for SARS-CoV-2, the virus responsible for the COVID-19 pandemic. MLE is a powerful tool for reconstructing the evolutionary history of any group of organisms, including coronaviruses [28].

Phylogenetic Analysis

This step involves using the evolutionary model and the multiple sequence alignments to infer the phylogenetic relationships among the selected organisms or gene family.

Phylogenetic analysis is a method of analyzing the evolutionary history of organisms. It is used to determine the relationships between different species and to understand the evolutionary processes that have occurred over time. In the case of SARS-CoV-2, phylogenetic analysis is used to construct phylogenetic trees which are diagrams that show the evolutionary relationships between different strains of the virus. Phylogenetic analysis involves collecting data from multiple sources, such as DNA sequences, protein sequences, and gene expression data. This data is then used to create a phylogenetic tree, which is a diagram that shows the evolutionary history of the virus. The phylogenetic tree is then used to compare different strains of the virus and to identify the different lineages of the virus. This information can be used to understand the spread of the virus and to identify which strains may be more dangerous than others. It can also be used to develop treatments and vaccines for SARS-CoV-2 [29].

Tree construction

This step involves constructing a phylogenetic tree based on the inferred relationships from the data obtained from the phylogenetic analysis [7]. The tree is a graphical representation of the evolutionary relationships among the organisms or gene family being studied. Phylogenetic tree construction is a process used to create a visual representation of the evolutionary relationships among different species. This is done by analyzing the genetic sequences of each species and creating a diagram that shows how they are related. The tree is constructed by grouping together species that share more similar genetic sequences, with the most closely

related species being grouped together at the tips of the branches. This tree can then be used to study evolutionary processes, such as speciation and adaptation [8] [30].

The tree can also be used to identify the common ancestor of a group of species, and can provide insight into the history and evolution of life on Earth. And then assessing how well the tree matches the original tree. The accuracy of the tree is then assessed using a measure of the tree's robustness, such as bootstrap support. Bootstrapping is a method used to evaluate the accuracy of a phylogenetic tree. It involves randomly sampling the data multiple times to create multiple trees, and then comparing the results. This allows researchers to see how reliable the tree is and determine whether it is accurate enough to make conclusions from. Bootstrapping can also be used to measure the confidence of a particular branch or node in the tree. The higher the bootstrap score, the more reliable the branch or node is considered to be [31]. After the process of Multiple Sequence Aligning as shown in Figure 4:

1. NC 048216.1 |NL63-related bat coronavirus strain BtKYNL63-9b complete genome
2. NC 046964.1 |Alphacoronavirus Bat-CoV/P.kuhlii/Italy/3388-19/2015 complete genome
3. NC 038861.1 |Transmissible gastroenteritis virus complete genome genomic RNA
4. NC 034972.1 |Coronavirus AcCoV-JC34 complete genome
5. NC 032730.1 |Lucheng Rn rat coronavirus isolate Lucheng-19 complete genome
6. NC 032107.1 |NL63-related bat coronavirus strain BtKYNL63-9a complete genome
7. NC 030292.1 |Ferret coronavirus isolate FRCoV-NL-2010 complete genome
8. NC 028752.1 |Camel alphacoronavirus isolate camel/Riyadh/Ry141/2015 complete genome
9. NC 028606.1 |Swine enteric coronavirus strain Italy/213366/2009 complete genome
10. NC 028811.1 |BtMr-AlphaCoV/SAX2011 complete genome
11. NC 028814.1 |BtRf-AlphaCoV/HuB2013 complete genome
12. NC 028824.1 |BtRf-AlphaCoV/YN2012 complete genome
13. NC 028833.1 |BtNv-AlphaCoV/SC2013 complete genome
14. NC 023760.1 |Mink coronavirus strain WD1127 complete genome
15. LC756668.1 |Human coronavirus NL63 HCoV/NL63 Fukushima 0230 2016 RNA nearly complete genom
16. ON461751.1 |Human coronavirus 229E isolate 229E 1 partial genome
17. ON461752.1 |Human coronavirus 229E isolate 229E 2 1 partial genome
18. NC 045512.2 |Severe acute respiratory syndrome coronavirus 2 isolate Wuhan-Hu-1 complete genom
(a)
1. NC 048217.1 |Murine hepatitis virus strain A59 complete genome
2. NC 045512.2 |Severe acute respiratory syndrome coronavirus 2 isolate Wuhan-Hu-1 complete genome
3. NC 006213.1 |Human coronavirus OC43 strain ATCC VR-759 complete genome
4. NC 038294.1 |Betacoronavirus England 1 isolate H123990006 complete genome
5. NC 039207.1 |Betacoronavirus Erinaceus/VMC/DEU/2012 isolate ErinaceusCoV/2012-174/GER/2012 complete genome
6. NC 030886.1 |Rousettus bat coronavirus isolate GCCDC1 356 complete genome
7. NC 026011.1 |Betacoronavirus HKU24 strain HKU24-R05005I complete genome
8. NC 025217.1 |Bat Hp-betacoronavirus/Zhejiang2013 complete genome
9. AC 000192.1 |Murine hepatitis virus strain JHM complete genome
10. NC 019843.3 |Middle East respiratory syndrome-related coronavirus isolate HCoV-EMC/2012 complete genome
11. NC 017083.1 |Rabbit coronavirus HKU14 complete genome
12. NC 014470.1 |Bat coronavirus BM48-31/BGR/2008 complete genome
13. NC 012936.1 |Rat coronavirus Parker complete genome
14. NC 009019.1 |Tylonycteris bat coronavirus HKU4 complete genome
15. NC 009020.1 |Pipistrellus bat coronavirus HKU5 complete genome
16. NC 009021.1 |Rousettus bat coronavirus HKU9 complete genome
17. NC 006577.2 |Human coronavirus HKU1 complete genome
18. NC 004718.3 |SARS coronavirus Tor2 complete genome
19. NC 003045.1 |Bovine coronavirus isolate BCoV-ENT complete genome
20. NC 001846.1 |Murine hepatitis virus strain MHV-A59 C12 mutant complete genome
21. OQ504244.1 |Severe acute respiratory syndrome coronavirus 2 isolate SARS-CoV-2/human/JPN/HiroFH291/2022 complete genom
22. OQ504245.1 |Severe acute respiratory syndrome coronavirus 2 isolate SARS-CoV-2/human/JPN/HiroFH376/2022 complete genom
23. OQ504246.1 |Severe acute respiratory syndrome coronavirus 2 isolate SARS-CoV-2/human/JPN/HiroFH377/2022 complete genom
24. OQ504247.1 |Severe acute respiratory syndrome coronavirus 2 isolate SARS-CoV-2/human/JPN/HiroFH378/2022 complete genom
25. OQ504248.1 |Severe acute respiratory syndrome coronavirus 2 isolate SARS-CoV-2/human/JPN/HiroFH380/2022 complete genom
26. OQ504274.1 |Severe acute respiratory syndrome coronavirus 2 isolate SARS-CoV-2/human/JPN/HiroC429/2022 complete genome
27. OQ504275.1 |Severe acute respiratory syndrome coronavirus 2 isolate SARS-CoV-2/human/JPN/HiroC430/2022 complete genome
28. OQ504276.1 |Severe acute respiratory syndrome coronavirus 2 isolate SARS-CoV-2/human/JPN/HiroC431/2022 complete genome
(b)

Figure.4. (a) Comparing Alpha coronavirus sequences to NC_045512.2 (the SARS-CoV-2 reference sequence) using a multiple sequence alignment (MSA). (b) Multiple sequence alignment of Beta

A phylogenetic tree was built for two families of Coronavirus, alpha and beta, with bootstraps as shown in Figures 5, 6.

Evaluation of phylogenetic tree

This step involves assessing the accuracy of the phylogenetic tree by comparing it to other trees that were constructed using different methods and assessing the support for each branch. This allows for the identification of any potential errors in the tree [32].

NC 048216.1 |NL63-related bat coronavirus strain BtKYNL63-9b complete genome
NC 032107.1 |NL63-related bat coronavirus strain BtKYNL63-9a complete genome
LC756668.1 |Human coronavirus NL63 HCoVNL63 Fukushima O230 2018 RNA nearly complete genome
NC 028752.1 |Camel alphacoronavirus isolate camel/Riyadh/Ry141/2015 complete genome
ON461751.1 |Human coronavirus 229E isolate 229E 1 partial genome
ON461752.1 |Human coronavirus 229E isolate 229E 2 1 partial genome
NC 046964.1 |Alphacoronavirus Bat-CoV/P.kuhlii/Italy/3398-19/2015 complete genome
NC 028833.1 |BtNv-AlphaCoV/SC2013 complete genome
NC 028811.1 |BtMr-AlphaCoV/SAX2011 complete genome
NC 028814.1 |BtRf-AlphaCoV/HuB2013 complete genome
NC 028824.1 |BtRf-AlphaCoV/YN2012 complete genome
NC 038861.1 |Transmissible gastroenteritis virus complete genome genomic RNA
NC 028806.1 |Swine enteric coronavirus strain Italy/213306/2009 complete genome
NC 030292.1 |Ferret coronavirus isolate FRCoV-NL-2010 complete genome
NC 023760.1 |Mink coronavirus strain WD1127 complete genome
NC 034972.1 |Coronavirus AcCoV-JC34 complete genome
NC 032730.1 |Lucheng Rn rat coronavirus isolate Lucheng-19 complete genome
NC 045512.2 Severe acute respiratory syndrome coronavirus 2 isolate Wuhan-Hu-1 complete genome

Figure.5. The Maximum Likelihood approach with bootstraps, a phylogenetic tree of (18) Alpha coronaviruses was created, with the SARS-CoV-2 RefSeq.

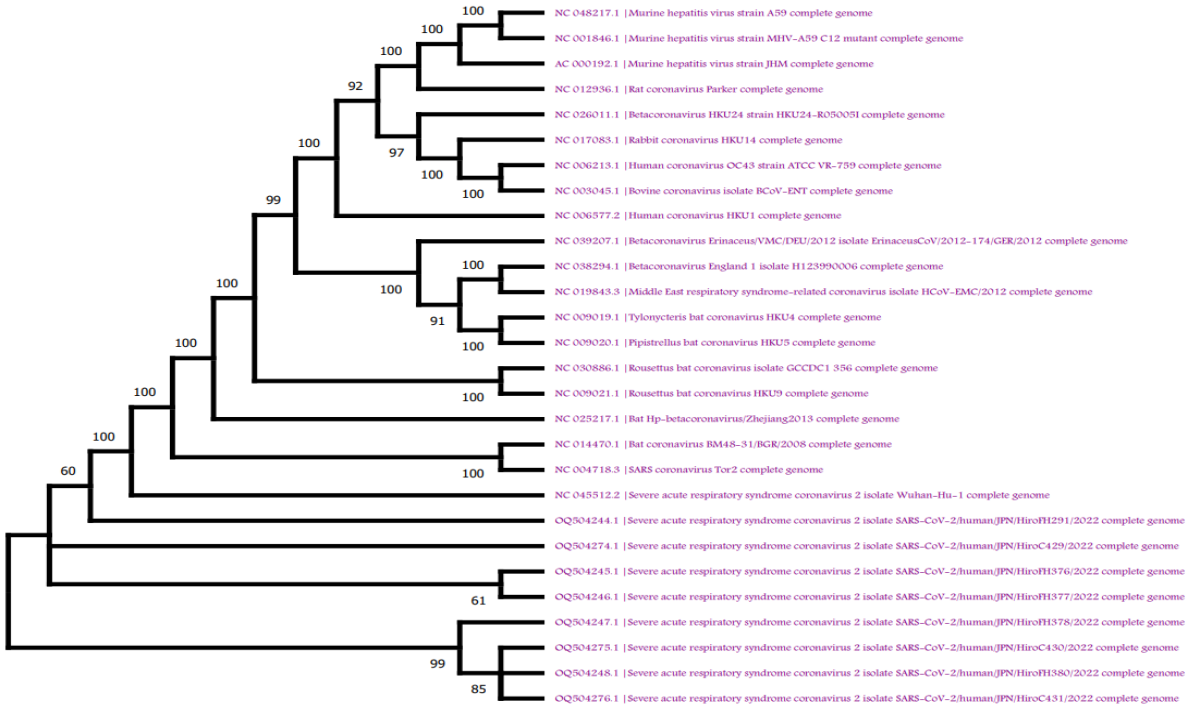
NC 048217.1 |Murine hepatitis virus strain A59 complete genome
NC 001846.1 |Murine hepatitis virus strain MHV-A59 C12 mutant complete genome
AC 000192.1 |Murine hepatitis virus strain JHM complete genome
NC 012936.1 |Rat coronavirus Parker complete genome
NC 026011.1 |Betacoronavirus HKU24 strain HKU24-R05005I complete genome
NC 017083.1 |Rabbit coronavirus HKU14 complete genome
NC 006213.1 |Human coronavirus OC43 strain ATCC VR-759 complete genome
NC 003045.1 |Bovine coronavirus isolate BCoV-ENT complete genome
NC 006577.2 |Human coronavirus HKU1 complete genome
NC 039207.1 |Betacoronavirus Erinaceus/VMC/DEU/2012 isolate ErinaceusCoV/2012-174/GER/2012 complete genome
NC 038294.1 |Betacoronavirus England 1 isolate H123990006 complete genome
NC 019843.3 |Middle East respiratory syndrome-related coronavirus isolate HCoV-EMC/2012 complete genome
NC 009019.1 |Tylonycteris bat coronavirus HKU4 complete genome
NC 009020.1 |Pipistrellus bat coronavirus HKU5 complete genome
NC 030886.1 |Rousettus bat coronavirus isolate GCCDC1 356 complete genome
NC 009021.1 |Rousettus bat coronavirus HKU9 complete genome
NC 025217.1 |Bat Hp-betacoronavirus/Zhejiang2013 complete genome
NC 014470.1 |Bat coronavirus BM48-31/BGR/2008 complete genome
NC 004718.3 |SARS coronavirus Tor2 complete genome
NC 045512.2 |Severe acute respiratory syndrome coronavirus 2 isolate Wuhan-Hu-1 complete genome
OQ504244.1 |Severe acute respiratory syndrome coronavirus 2 isolate SARS-CoV-2/human/JPN/HiroFH291/2022 complete genome
OQ504274.1 |Severe acute respiratory syndrome coronavirus 2 isolate SARS-CoV-2/human/JPN/HiroC429/2022 complete genome
OQ504245.1 |Severe acute respiratory syndrome coronavirus 2 isolate SARS-CoV-2/human/JPN/HiroFH376/2022 complete genome
OQ504246.1 |Severe acute respiratory syndrome coronavirus 2 isolate SARS-CoV-2/human/JPN/HiroFH377/2022 complete genome
OQ504247.1 |Severe acute respiratory syndrome coronavirus 2 isolate SARS-CoV-2/human/JPN/HiroFH378/2022 complete genome
OQ504275.1 |Severe acute respiratory syndrome coronavirus 2 isolate SARS-CoV-2/human/JPN/HiroC430/2022 complete genome
OQ504248.1 |Severe acute respiratory syndrome coronavirus 2 isolate SARS-CoV-2/human/JPN/HiroFH380/2022 complete genome
OQ504276.1 |Severe acute respiratory syndrome coronavirus 2 isolate SARS-CoV-2/human/JPN/HiroC431/2022 complete genome

Figure.6. The Maximum Likelihood approach with bootstraps, a phylogenetic tree of (28) Beta coronaviruses was created, with the SARS-CoV-2 RefSeq.

The evaluation of a phylogenetic tree depends on the accuracy of the data used to build the tree. The accuracy of the data is determined by factors such as the quality of the data sources, the reliability of the analysis methods used, and the accuracy of the inference techniques used. If the data used is of high quality, the phylogenetic tree will be more accurate .

The accuracy of the branch lengths can be evaluated by looking at how well the tree reflects the actual evolutionary history of the species. The certainty of the conclusions drawn from the tree can be evaluated by considering the strength of the evidence that supports the tree [33] [34].

Results of Tree construction

The Results of this research paper covers the results of the study and the interpretation of these outcomes. We used the maximum likelihood estimate (MLE) approach to generate three phylogenetic trees based on complete genome sequences. The first tree (Figure 5) was for Alpha coronavirus species only, the second (Figure 6) was for Beta coronavirus species, including SARS-COV-2 Species, and the third (Figure 7) was constructed by merging the Alpha and Beta phylogenetic trees.

Conclusion

In this paper, we used the Maximum Likelihood Estimate (MLE) approach to construct a phylogenetic tree of Alpha and Beta coronavirus to illustrate the evolutionary ties between SARS-CoV-2 and (Alpha and Beta families) of coronavirus. This resulted in accurate and dependable phylogenetic tree. It is likely that different variants of the COVID-19 virus will emerge quickly, so the different strains of SARS-CoV-2 must be kept an eye on. Utilizing the phylogenetic tree and the Maximum Likelihood Estimation (MLE) approach appears to be a successful and advantageous technique for tracking the advancement of SARS-CoV-2's lineage, since the tree generated is accurate.

Future Works

In the future, we plan to build a phylogenetic tree of SARS-CoV-2 to demonstrate the evolutionary links between Alpha, Beta, Gamma, and Delta coronaviruses, which will uncover various SARS-CoV-2 mutations and single nucleotide polymorphisms (SNPs). To contain and treat COVID-19 and stop another outbreak, it is imperative to continually observe the variety and evolution of SARS-CoV-2.

Acknowledgments

The authors would like to express their appreciation to Mustansiriyah University (www.uomustansiriyah.edu.iq) Baghdad – Iraq and Diyala University for providing assistance in this paper.

References

[1] M. F. Boni, Nature Microbiology, vol. VOL 5, pp. 1408-1417, November 2020.

[2] R. E. Anirudh Sridhar, "LEVERAGING A MULTIPLE-STRAIN MODEL WITH MUTATIONS IN ANALYZING THE SPREAD OF COVID-19," IEEE, pp. 8163-8167, February 24,2023.

[3] J. E. Lemieux, "Phylogenetic analysis of SARS-CoV-2 in Boston highlights the impact of superspreading events," Science 371, vol. eabe3261, pp. 1-9, 5 February 2021.

[4] S. SRIVASTAVA, "SARS-CoV-2 genomics: An Indian perspective on sequencing viral variants," Journal of Biosciences-Indian Academy of Sciences, vol. 46, no. 22, pp. 1-14, 2021.

[5] G. Maged N. Kamel Boulos, "Geographical tracking and mapping of coronavirus disease COVID-19/ (SARS-CoV-2) epidemic and associated events around the world: how 21st century GIS technologies are supporting the global fght against outbreaks and epidemics," International Journal of Health Geographics, vol. 19, no. 8, pp. 1-12, (2020).

[6] M. Ghada Younis ABDULRAHMAN, "IMMUNOLOGICAL RESPONSE TO SARS-Cov 2 AND IMMUNOPATHOLOGY OF COVID-19 INFECTION," International Journal of Applied Sciences and Technology , vol. 2717, no. 8234, pp. 33-44, /2022.

[7] B. Al-Nuaimi, Ancestral Reconstruction and Investigations of Genomics Recombination on Chloroplasts Genomes, France: UNIVERSITÉ F R A N C H E - COMTÉ, Submitted on 11 Apr 2019.

[8] B. Alkindy, Combining approaches for predicting genomic evolution, France: UNIVERSITÉ F R A N C H E - COMTÉ, Submitted on 6 Jan 2017.

[9] A.-N. Alaa Khudair Abbas Al-Khafaji, "Phylogenetic Tree Construction to Reveal the Detailed Evolution of SARS-CoV-2," JOURNAL OF ALGEBRAIC STATISTICS, vol. 13, no. 2, pp. 538 - 549, 2022.

[10] Mohammed Uddin, "SARS-CoV-2/COVID-19: Viral Genomics, Epidemiology, Vaccines, and Therapeutic Interventions," Viruses, vol. 12, no. 526, pp. 1-18, 2020,.

[11] Lu Lu, "Adaptation, spread and transmission of SARS-CoV-2 in farmed minks and associated humans in the Netherlands," NATURE COMMUNICATIONS, vol. 12, no. 6802, pp. 1-12, 2021.

[12] F. C. Bashar Talib Al-Nuaimi, "Relation between Gene Content and Taxonomy in Chloroplasts," arXiv:1609.06055v1 [q-bio.GN], September 2016.

[13] A. N. Christophe Guyeux, "On the Ability to Reconstruct Ancestral Genomes from Mycobacterium Genus," Conference: International Conference on Bioinformatics and Biomedical Engineering, April 2017.

[14] "NCBI Virus" (https://www.ncbi.nlm.nih.gov/labs/virus/vssi/#/virus) Accessed on 2023.," [Online].

[15] M. C. Pryavahiny Kichenaradja, "ISbrowser: an extension of ISfinder for visualizing insertion sequences in prokaryotic genomes," Nucleic Acids

Research, vol. 38, no. Database, pp. 62-68, 2010.

[16] D.-S. DOUGLAS R. SMITH, "Complete Genome Sequence of Methanobacterium thermoautotrophicum DH: Functional Analysis and Comparative Genomics," JOURNAL OF BACTERIOLOGY, vol. 179, no. 1997, pp. 7135-7155, Nov. 1997.

[17] A. Coghlan, A Little Book of R For Bioinformatics, Release 0.1, Aug 18, 2017.

[18] D. J. Zwickl, "GENETIC ALGORITHM APPROACHES FOR THE PHYLOGENETIC ANALYSIS OF LARGE BIOLOGICAL SEQUENCE DATASETS UNDER THE MAXIMUM LIKELIHOOD CRITERIONThe," University of Texas at Austin, 2016.

[19] M. G. Nicolas Galtier, "Inferring Pattern and Process: Maximum-Likelihood Implementation of a Nonhomogeneous Model of DNA Sequence Evolution for Phylogenetic Analysis," Molecular Biology and Evolution, no. 0737-4038, pp. 871-879, 1998.

[20] D. Vijaykrishna, "Evolutionary Insights into the Ecology of Coronaviruses," Virology, vol. 81, no. 8, pp. 4012-4020, 15 April 2007.

[21] T. V. AMY BOUCK, "The molecular ecologist's guide to expressed sequence tags," Molecular Ecology, vol. 16, p. 907–924, 2007.

[22] T.-J. Chang, "Genomic analysis and comparative multiple sequences of SARS-CoV2," pp. 537-543, 2020.

[23] H. P. Julian M. Catchen, "Automated identification of conserved synteny after whole-genome duplication," Cold Spring Harbor Laboratory Press, pp. 1497-1505, February 24, 2023.

[24] P. B. Benoit Morel, "Phylogenetic Analysis of SARS-CoV-2 Data Is Difficult," p. 1777–1791, 2020.

[25] R. Gentleman, R Programming for Bioinformatics, Computer Science and Data Analysis Series, 2009.

[26] D. Rochman, "Ongoing Adaptive Evolution and Globalization of Sars-Cov-2," October 16, 2020.

[27] N. Kumar, "Evolutionary Signatures Governing the Codon Usage Bias in Coronaviruses and Their Implications for Viruses Infecting Various Bat Species," Viruses, vol. 13, no. 1847, pp. 1-18, 2021.

[28] X. L. Ming-Ze Yin, "Parameter estimation of the incubation period of COVID-19 based on the doubly interval-censored data model," Nonlinear Dyn, no. 106, p. 1347–1358, 2021.

[29] J. J. Sikkens, "Serologic Surveillance and Phylogenetic Analysis of SARS-CoV-2 Infection Among Hospital Health Care Workers," JAMA Network Open, pp. 1-13, 2021.

[30] J. D. Ali .F. HUSSEIN, "INCIDENCE OF COVID 19 IN RELATION TO ENVIRONMENTAL TEMPERATURE AT AL- BASRA," International Journal of Applied Sciences and Technology, no. 2717-8234, pp. 111-118, 2022.

[31] Tingting Li, "Phylogenetic supertree reveals detailed evolution of SARS-CoV-2," Scientifc Reports, 2020.

[32] M. S. Marini, "Regaining perspective on SARS-CoV-2 molecular tracing and its implications," medRxiv, March 20, 2020.

[33] K. N. Alireza Tabibzadeh, "Evolutionary study of COVID-19, severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) as an emerging coronavirus: Phylogenetic analysis and literature review," Vet Med Sci, p. 559–571, 2021.

[34] L. Velazquez-Salinas, "Positive Selection of ORF1ab, ORF3a, and ORF8 Genes Drives the Early Evolutionary Trends of SARS-CoV-2 During the 2020 COVID-19 Pandemic," Frontiers in Microbiology, vol. 11, pp. 1-13, 2020.